MarketingManual_GroupAssignment

Tatiane Dutabruno - Charlotte Gallet - Harikrishnan Gopalakrishnan

18/12/2021

Gambling Marketing Data Mart

The objective of this marketing manual is to understand and capture the important marketing metrics for a betting company at a customer level.

In the following manual, you will find the presentation of the data available, after which will be an explanation regarding new variables and it will end with some graphs to explain the findings.

Presentation of the Datasets

While working on this online gambling company, three data sets were used. The first data set, called Demographics, contained information regarding personal information on the clients. This type of information would be gender, country of residence, spoken language, which allows to have insights on some categories of people, in ways that could improve targeting of the clients. On the other hand, this data set also contained information linked to the client and their relationship with the company. Some features would be the registration date, the first active date (or the first date where they actively used the website). Not all information could be used and analyzed, due to privacy concerns, but most of the information that has been gathered allowed to understand and create deeper understandings of the clients.

The second data set we had access to, named User Daily Aggregation, focuses on the products that are available on the gambling website. The table consists of features like the date, the customer and the product associated to the betting that was done. The betting information seems to be the most important as it shows some numerical information, like the amount of winnings or the stakes. That way, it is possible to have better insights when looking at groups of products and the outcome for each client or group of client. The specificity of this data set is that it reunites all type of products from sports bets to games or casinos. It is possible to see that one of the product is dedicated to poker games, however, no information regarding this type of product is collected in the data set. This can be explained by the next data set.

Indeed, the third and last data set, Poker Chip Conversions, which contains all the information regarding only transactions with poker chips. The features that have been used are linked to the amount of each transaction, the date of the transaction and the type of transaction. There are two different types of transaction, the first one being a "buy" transaction and the second a "sell". From our understanding, a "buy" transaction reflects to the number of poker chips has bought, in order to bet in a game; while a "sell" transaction would be the amount the client has gained and is willing to transform from chips into money.

Now that we know what type of information we have and what insights we could get from it, it was necessary to create one single data set which unites all the information. In order to do so, some data preparation was needed. The main data set requires to have one row per customer with all the information necessary. Also, we are only looking at clients who have been registered between the 1st and the 27th of February 2005 and the transactions needed to have been executed between the 1st of February 2005 until the 30th of September 2005. Data cleaning, regarding missing values especially in the first active dates, needed to be removed to have a clearer view and no errors in the data set. Finally, the critical aspect of this data mart was to only use the records that took place after or on the first the first pay-in date of the customer.

Data Preparation

Hence all the constraints, some changes had to be made within the data sets. In the following part, explanation regarding the creation of new variables will be given as well as some adaptation to be able to join all the data sets as one whole data set.

Concerning the Demographics data set, no features were removed but some were added. As we are taking a marketing perspective on the gambling website and their clients, we though it would be necessary to have a length of relationship variable calculated based on the date of registration up until the 30th of September 2005, which is the last date of transaction. The second aspect that was interesting to calculate was the difference in time between the registration date and the first date of purchase, as some clients might have bet directly after registering or if some time passed between the two moments. This data set is now ready to be used in a bigger data mart.

The next data set required a little more preparation. Indeed, after some basic cleaning, it was time to create new variables. The first variables rely on each client. That is looking at the total, average, minimum and maximum of each features which are the stakes, the winnings and the bets. Once these features had been created it was possible to further the information with variables like the frequency of days played, the Gross Gaming Revenue of the user(Total Winnings - Total Stakes), the Average Gaming Revenue of the user((Total Winnings - Total Stakes)/Frequency). Some information could also be gathered regarding the company, taking into consideration that the stakes would create some revenue for the company while any winnings would be a loss. That is how the company's Gross Gaming Revenue and the Average Gaming Revenue have been calculated (Total Stakes - Total winning or (Total Stakes - Total Winnings)/Frequency).
The information above looks at each client individually, however it was also necessary to group the information per product, as to have a bigger picture of the information, rather than specific to each client. That is why after some data manipulation to group all the information related to each product, it was possible to find information concerning the frequency per product, the total bets per product and finally the company's Gross Gaming Revenue generated per product. Once all these steps were accomplished, it was possible to go to the next data set to finish preparing them all for the final merge.

It was now possible to do the same type of preparation with the Poker Chip Conversion data set. That is, removing all the unnecessary information, transforming some variables to adapt to the correct formats. From there, variables like the total, average, minimum and maximum of buys, sells and number of transactions could be calculated. With that ensued the creation of variables like the Gross Gaming Revenue of the consumer and of the company, paired with its average. This time though, the information would be specifically regarding the poker bets.

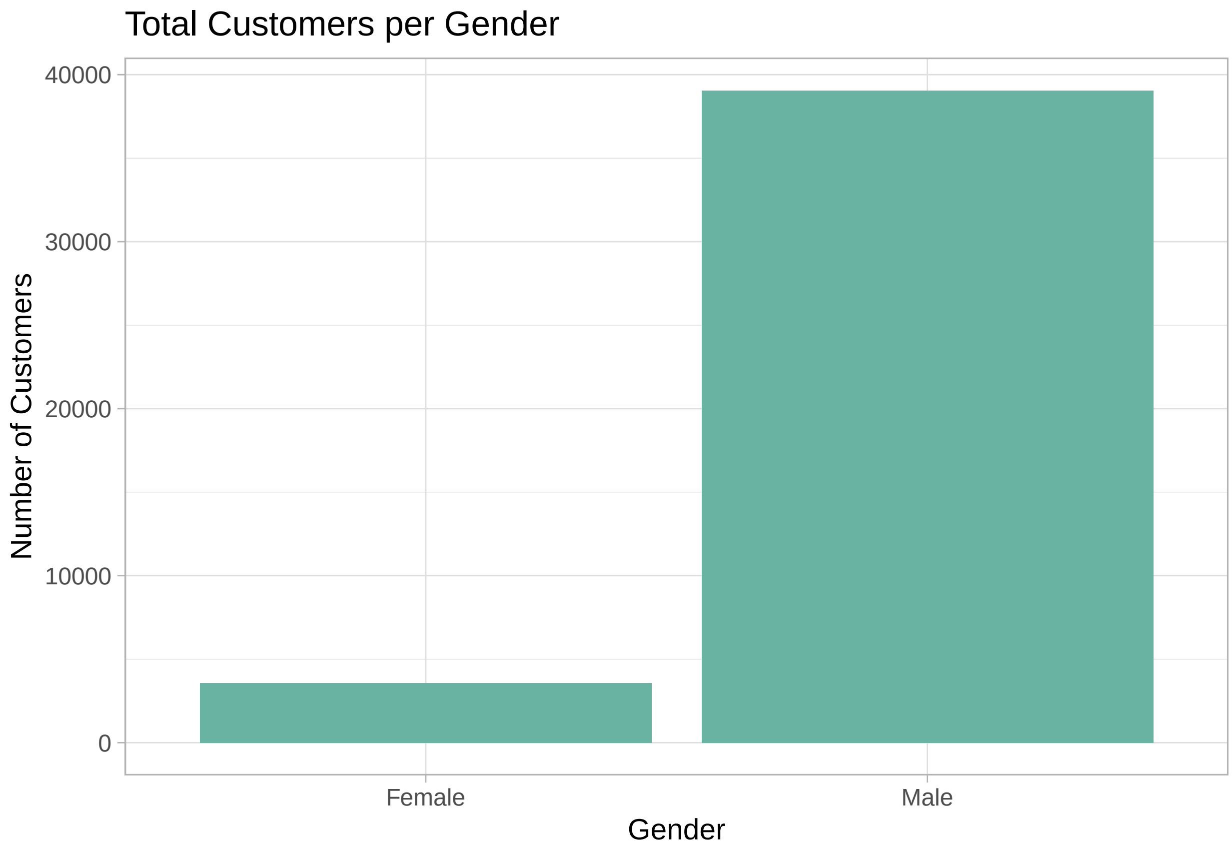
Finally, all these tables were merged together to create one bigger data set. From there, more variables could be found, as it was possible to get information from various tables together and find out new and more complete information. For instance, most of the variables regarding the frequency, the bets, the Gross Gaming Revenue, the recency, the total spent have been aggregated between the information of each product and the information from the poker table. From these new variables it was possible to calculate the Recency Frequency and Monetary score as well as the stickiness, or whether a client was addicted or not.

Once the data mart was complete and all the information gathered was mixed and integrated with one another, it was possible to gather some insights on the clients and what they could offer to the company.

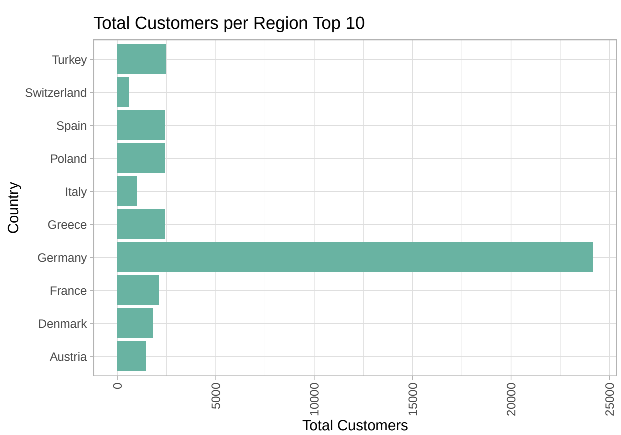
Marketing Insights

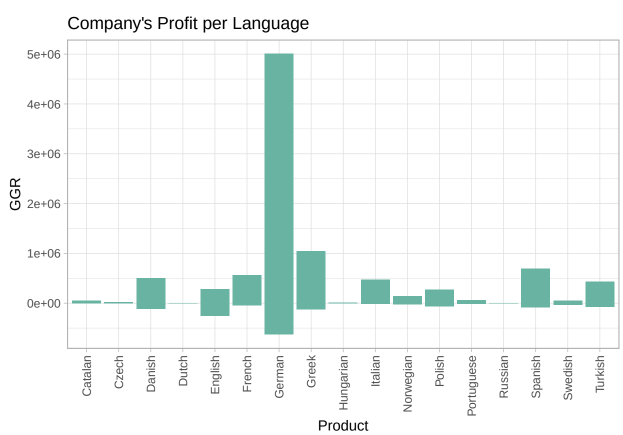
To gather some insights, we have made some graphs with the main aspects of the clients.

First, it is interesting to look at the difference of clients between gender. As it is possible to see in the below graph, the amount of female client seem to represent approximately 1% of the total clients. This shows to say that it would be more likely that men are more interested in betting on this website than women.

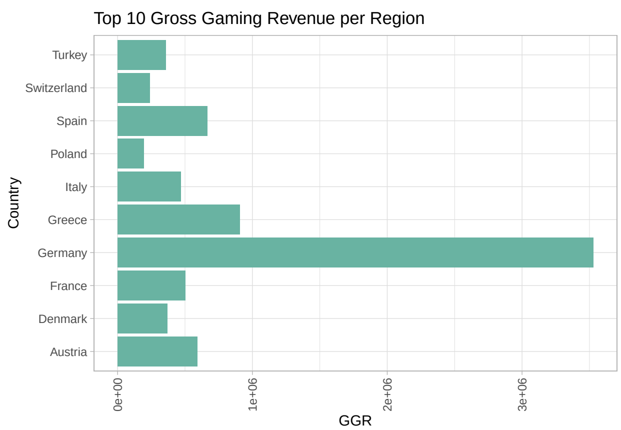


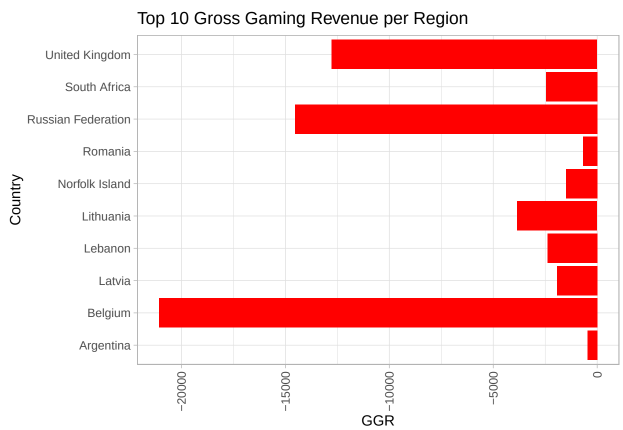
For some more demographic information, the following graph will show the amount of customers per country of residence. As there are over 200 countries in the list, we have decided to select the top 10 countries, the 10 countries where there are the most clients. From this, it is possible to see that the largest amount of customers come from Germany. This will explain the graph concerning the highest profit according to language, which is where the clients speak German.



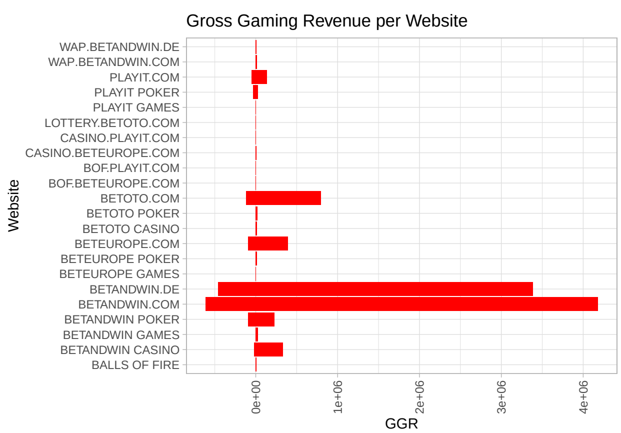


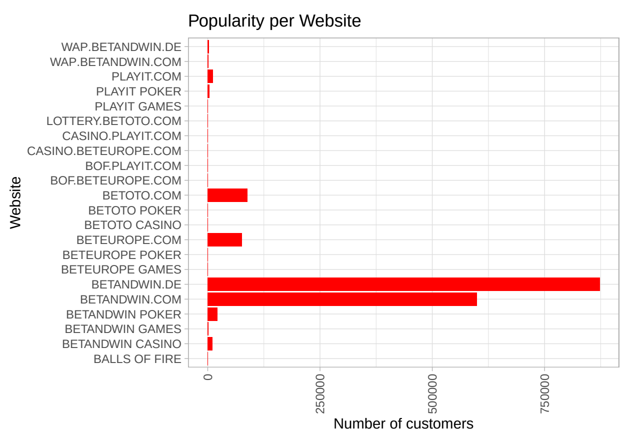
Now that some demographics have been explained, it will be interesting to look at the Gross Gaming Revenue of the company with regard to the top 10 countries where the revenue is the highest. This graph is on the left hand-side, once again Germany seems to be the country that brings the highest amount of revenue to the company. On the right hand-side, we can see a graph that represents that 10 countries that bring the least amount of revenue to the company. That is, clients from these countries tend to not contribute to the overall profit of the company.



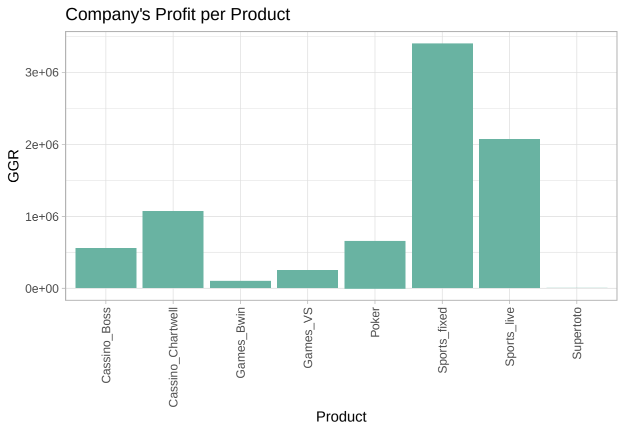


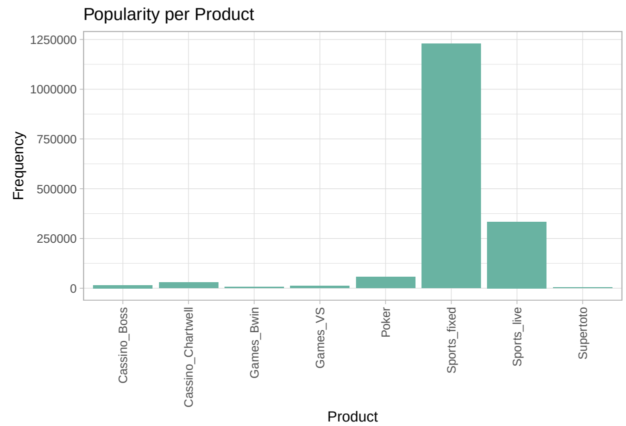
To further these insights, some information regarding which application or website generates the highest revenues and is therefore the most popular within the clients and add the biggest revenues to the company. It seems as though BetAndWin have the strongest position, both in terms of revenue and in terms of number of customers. We can see that the once bringing the most revenues to the company seems to be the website of the world (.com), whereas, the one where there are the most amount of clients is the German website (.de).



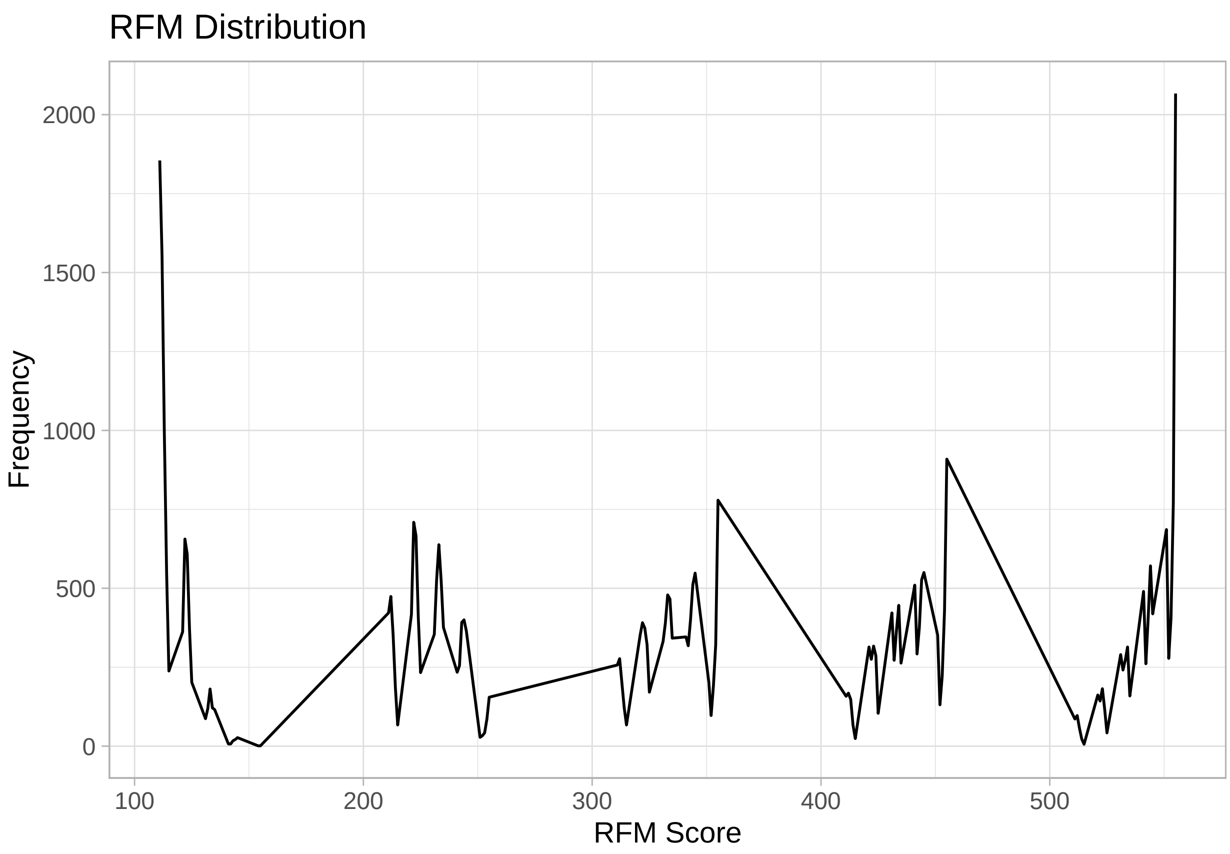


It is also interesting to take a look at the products and their contribution to the company. As it is possible to see in the following graphs, the sports category seem to be the most popular products as they generate the most revenue as well as are the most frequently used.



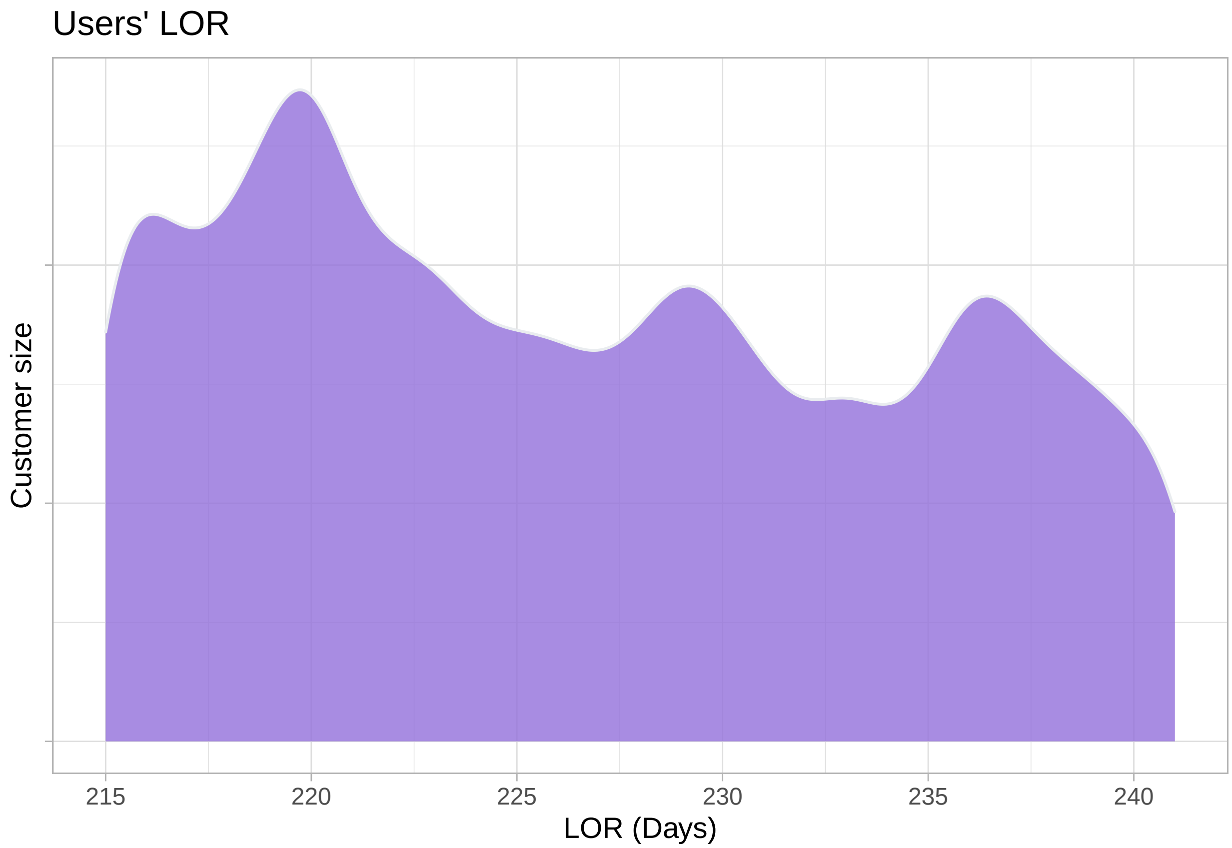


Looking at the RFM Score in terms of the frequency, there seems to be a relationship between frequency and the RFM score. As it is possible to see in the below graph, it seems as though there are two extremes. When the score is very low or very high, is where we can observe the highest frequency. Throughout the rest of the observations, we can see some peaks of frequency when the score is relatively high. It seems as though there is a sort of cyclical relationship between the frequency and the RFM score.

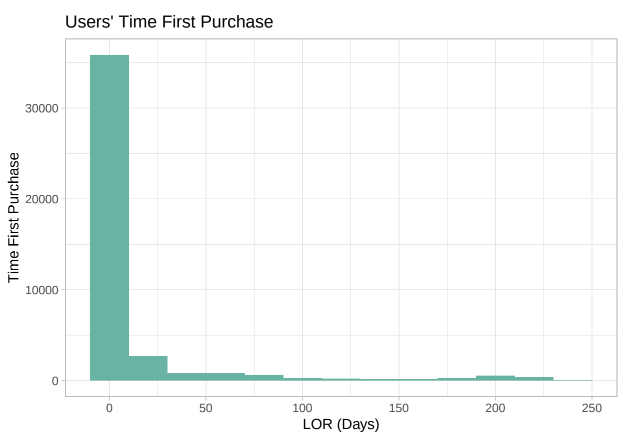


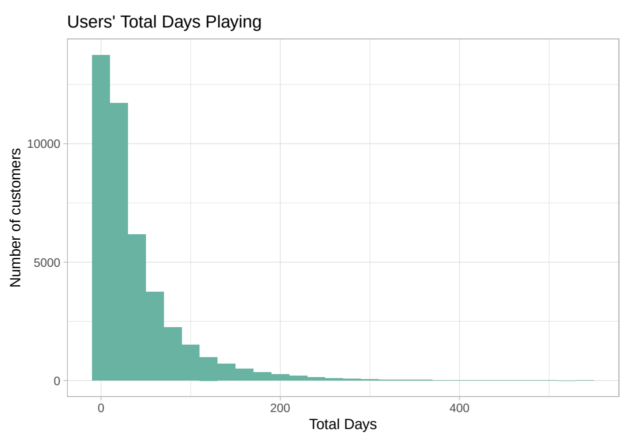
Now that we have seen some information regarding the scores and the contribution of the gamblers to the company, through different perspectives, below will be few graphs that will look at the characteristics of these clients.

For instance, looking at the amount of players and their habits throughout their relationship with the company. It looks like the longer a player has been with the company, the more he will bet. However, this needs to be nuanced, as it is possible to see a peak of customers at around 7 months of relationship with the company.

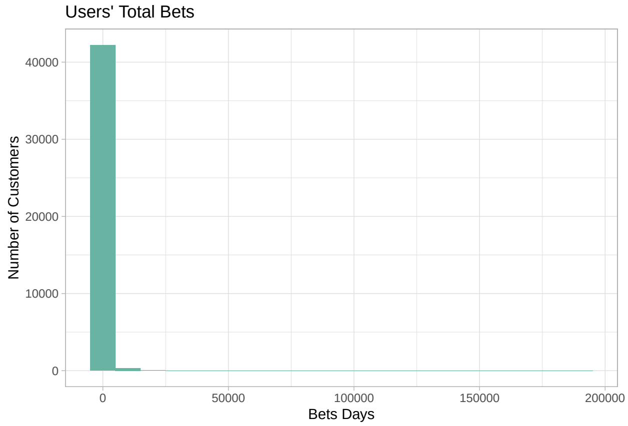


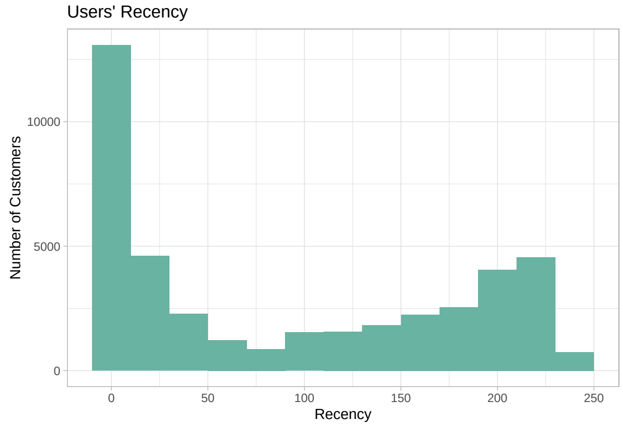
As we have added a feature which looks at the amount of days that have passed between the registration date and the first time the player has actually played, it is interesting to see this through a chart that could show some more insights. In the chart, we can see that the first time a gambler plays is often correlated with early stages of its relationship with the company. This information can also be looked at through a different perspective, that is to say, in the following graph, it is possible to understand that most of the customers will play during the first days of its relationship and registration with the company. Once a month has gone by, the amount of people playing decreases significantly.



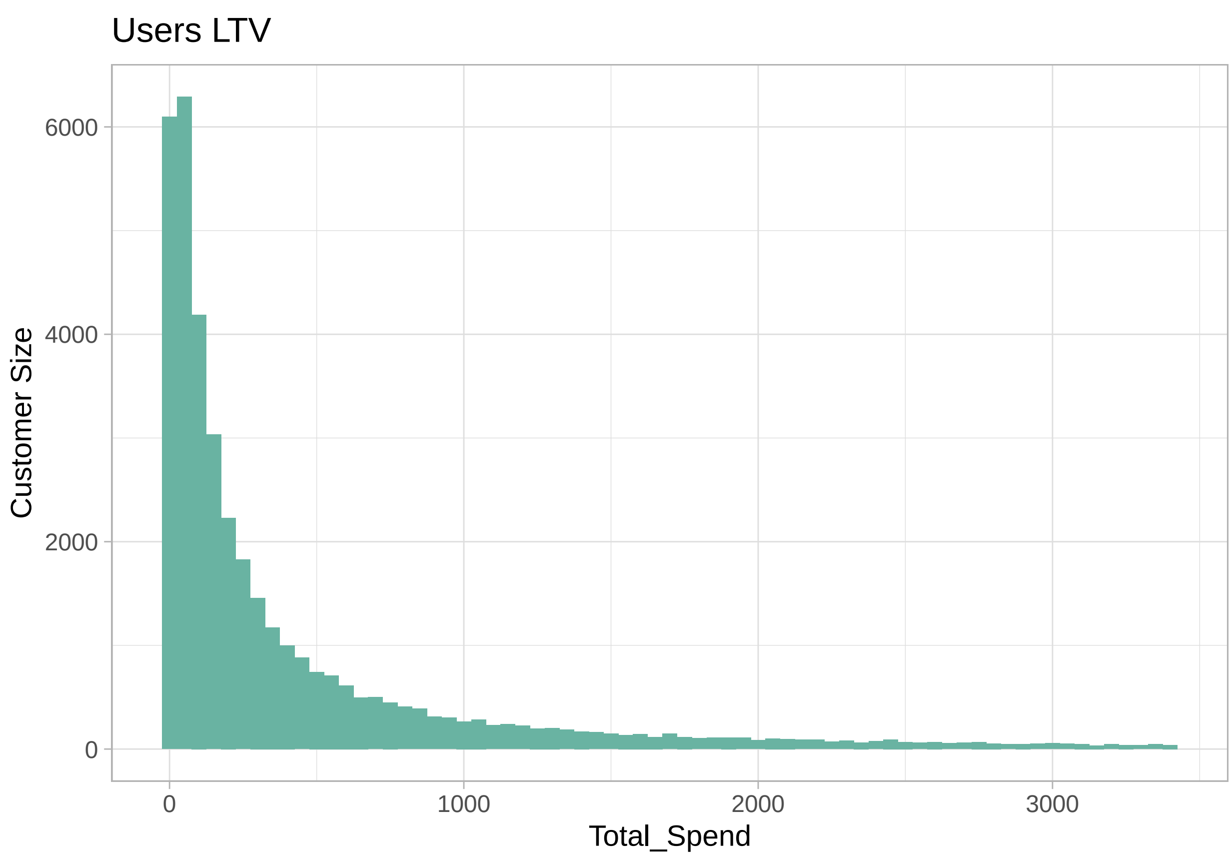


Furthermore, we can see that the total amount of gamblers playing often occur within the very beginning of their interest in gambling and the amount of days they spend gambling. This can be similar when looking at how recently a player has gambled. That is to say that the more recently a gambler has played the more he has played. Once again this can be nuanced as we can see that customers who have played a long time ago also seem to have played a lot. Maybe this can be explained by a high amount of bets at the beginning of a relationship and then not looking at the account and not placing any more bets onto the company anymore. In other words, what we see here is the churning rate, as the people who have last played more than 225 days ago are the ones who will not play anymore.

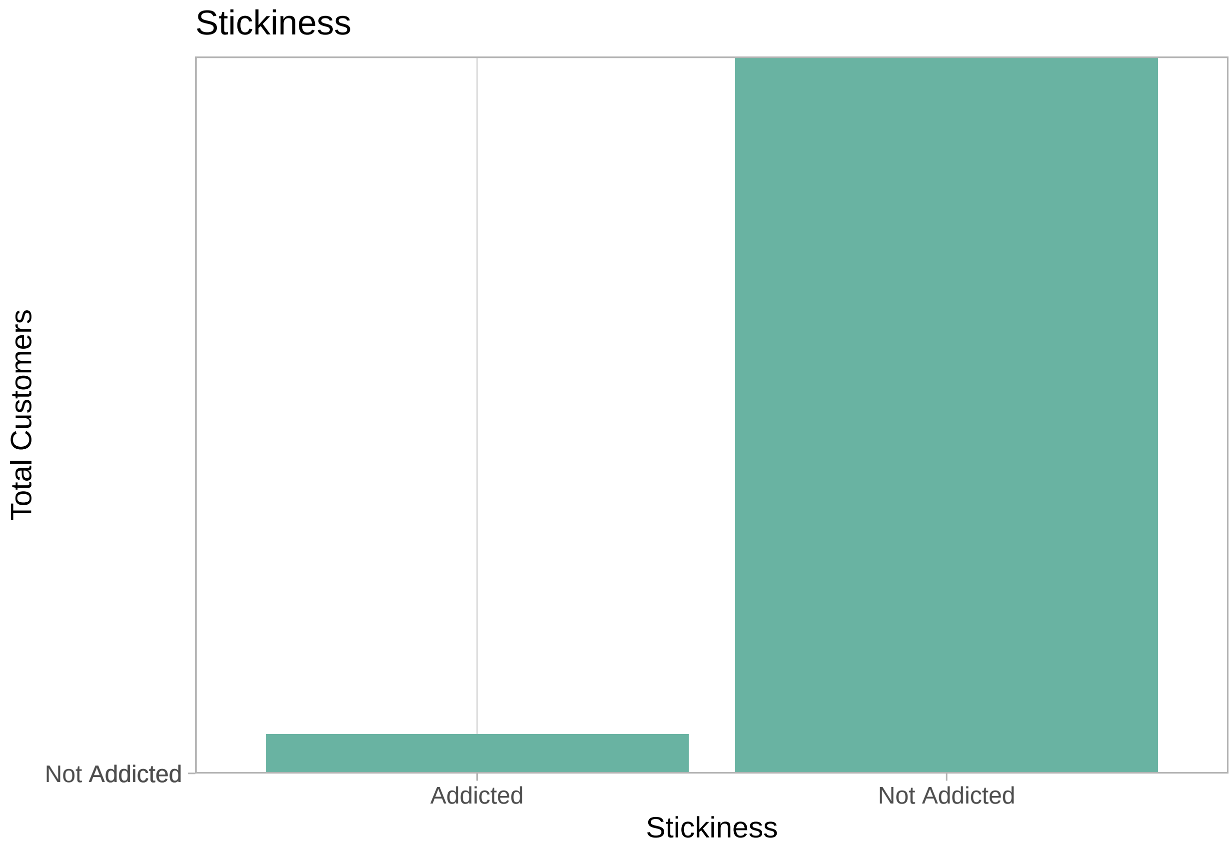




Now looking at some monetary aspect of the gamblers, it is possible to see, in the following graph, that the amount of gamblers playing can be perceived in terms of spending. Indeed, the more there are gamblers, the less they spend. That could be quite interesting as it might be the contrary, the more there are players, the more they will spend. As shown in the below graph, it is possible to see that this is not the case. An interesting insight that has been gained from this graph is also taking a look at the amount that the customers spend, only few have spent over 2,000 which shows the lifetime values of the customers. The conclusion to this insight shows that there are only few customers with high lifetime value.



Finally, as this is a gambling website it is important to track if there might be any issue regarding addiction and gamblers who might become too obsessed with gambling. As is possible to see in the below graph, it seems that only a small proportion of the overall clients are addicted while the biggest part of their customers are not addicted to gambling.



Conclusion

Thanks to this analysis, it is possible to have a better understanding of the company and its clients. We can say that the majority of the revenue and the gamblers are people who come from Germany, speak German and play on "BetandWin.de" website. The company's client are not very addicted gamblers who tend to spend and play as soon as they register and the frequency and amount spent diminishes with time passing by. These elements will help the company better focus its offerings, help target the appropriate people and make sure that the efforts in targeting people will pay off.

References

A. (2021, November 23). *16 metrics gaming app marketers must measure.* AppsFlyer. link

Academy, R. (2020). /rfm | Rsquared Academy.* Rfm.Rsquaredacademy. link

Bhalla, D. (2015). *R : Keep / Drop Columns from Data Frame.* ListenData. link

Blog, R. A. (2019, February 12). *RFM Analysis in R.* R-Bloggers. link

Delete or Drop rows in R with conditions. (2020, September 19). DataScience Made Simple. link

Garg, H. (2021, December 19). *Customer Segmentation using RFM analysis in R - Analytics Vidhya.* Medium. link

ggplot2 axis ticks : A guide to customize tick marks and labels - Easy Guides - Wiki - STHDA. (2020). STHDA. link

Holtz, Y. (2018). *Two Histograms with melt colors.* The R Graph Gallery. link

How to find the statistical mode? (2010, March 30). Stack Overflow. link

How to fix spaces in column names of a data.frame (remove spaces, inject dots)? (2012, May 21). Stack Overflow. link

How to remove warning messages in R Markdown document? (2017, July 30). Stack Overflow. link

Main Online Casino KPIs to Analyze and Improve. (2021). InTarget. link

Quick-R: Date Values. (2017). Quick-R by Datacamp. link

Shiny Dashboard. (2014). Shiny Dashboard. link

Using RFM to Identify Your Best Customers | Eight Leaves | Data Analytics & Software Development. (2011). Eight Leaves. link

Wickham, H. (2021). *Pivot data from wide to long — pivot_longer.* Tidyr. link

With min() in R return NA instead of Inf. (2018, January 19). Stack Overflow. link

Xie, Y. C. D. (2021, October 7). *R Markdown Cookbook.* Bookdown. link